Math 135A: HW1
Due Monday, January 13

1. Suppose A, B, and C are any three events in the probability space $(\Omega, \mathbb{P}, \mathcal{F})$. Prove that

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

2. In a 10 question true-false exam, find the probability that a student gets a grade of 70% or better by guessing. Answer the same question if the test has 30 true-false questions. Use your favorite software (MatLab, Maple, or Mathematica) to numerical evaluate your theoretical expressions. Hint: Binomial distribution.

3. The *Birthday Problem* asks what is the probability $P_{r,1}$ of finding at least one similar pair having the same birthday in a group of r individuals. In class we showed that

$$P_{r,1} = 1 - \frac{(365)_r}{365^r}$$

where $(n)_r$ is the Pochhammer symbol $(n)_r = n(n-1)\cdots(n-r+1)$. We found that for $r = 23$, $P_{23,1} \approx 1/2$.

   An extension of the birthday problems asks what is the probability $P_{r,2}$ of finding at least *two* matches. (Two matches occurs if there are two distinct pairs of common birthdays or with three people having the same birthday.) Write a computer program that simulates the birthday problem and estimate the number of people r so that that $P_{r,2} \approx 1/2$.

4. Each of 50 states has two senators. In a committee of 50 senators chosen at random, what is the probability of the event a given state is represented? Hint: First calculate the probability that the given state is *not* represented. Numerically evaluate your theoretical expressions.

5. **Background:** The *hypergeometric distribution*, $h(N, k, n, x)$, arises in the following way:

   Suppose we have N balls k of which are red and $N-k$ are blue. We draw a sample, without replacement, of n balls. Let X equal the number of red balls in our sample of size n. Let

$$C_{n,k} := \left(\begin{array}{c} n \\ k \end{array} \right) = \text{binomial coefficient}$$

   then[1]

$$\mathbb{P}(X = x) = h(N, k, n, x) = \frac{C_{k,x} C_{N-k,n-x}}{C_{N,n}} \tag{1}$$

[1] To see why (1) is true. First note the total number of all samples with no restrictions on the number of red or blue balls is $C_{N,n}$. The number of ways to choose x red balls is $C_{k,x}$. For each choice of red balls, there are $C_{N-k,n-x}$ ways to choose the blue balls. Thus there are $C_{k,x} \cdot C_{N-k,n-x}$ possible samples with x red balls and $k - x$ blue balls.

Exercise: From Feller, Vol. 1: "Suppose that 1000 fish caught in a lake are marked by red spots and released. After a while a new catch of 1000 fish is made, and it is found that 100 among them have red spots. What conclusions can be drawn concerning the number of fish in the lake?" It could be that there are only 1900 fish in the lake, but the probability of this is extremely small. A better estimate is to use the hypergeometric distribution where

$$\begin{aligned} N &= \text{unknown number of fish in the lake} \\ k &= 1000 \text{ fish with red spots (the red balls)} \\ n &= 1000\text{, sample size} \\ x &= 100 \text{ fish with red spots in sample} \end{aligned}$$

We now choose as our estimate for N that N that maximizes the probability $h(N, 1000, 1000, 100)$. Using your favorite software to find the estimate for N.